

THE GRADUATION
OF
EXPERIMENTAL DATA

BY

HENRY C. T. EGGERS

THE GRADUATION OF EXPERIMENTAL DATA

BY

HENRY C. T. EGGERS

*Assistant Professor of Drawing and Descriptive Geometry
University of Minnesota, Minneapolis, Minn.*



*A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy, in the
University of Michigan*

OFFICE OF THE EXECUTIVE SECRETARY
BELLEVUE COURT BUILDING
PHILADELPHIA

This paper is the literary property of the Institute and is not to be reprinted in whole or abstract except by permission and with due credit. Written discussion is invited and may be sent to the Executive Secretary, Frederic J. LeMaistre, 809 Bellevue Court Building, Philadelphia, Pa.



THE GRADUATION OF EXPERIMENTAL DATA

BY HENRY C. T. EGGERS*

CONTENTS

	PAGE
Introduction	I
1. The Problem of Graduation	I
2. Other Methods of Smoothing	2
(a) Graphical Methods: Sprague's Method	2
(b) Fitting Data to an Assumed Law	2
(c) Summation Formulae	3
3. The Principle of Areas	4
4. The Method of Graduation Presented in This Paper	4
Description of Method	5
5. Definitions	5
6. Explanation of the Method	6
7. Details of the Method	6
8. Calculations	10
9. Correction Factor	13
Illustrative Examples	18
10. The Cooling of a Vessel of Water	19
(a) Solution	19
(b) Comparison with the Solution by the Method of Least Squares	24
(c) Interpolation	25
11. The Reaction of Diphenyl-chlor-methan with Excess of Alcohol	28
(a) Pseudo Readings	28
(b) Smoothing of Rough Data	29
(c) Solution	30
Appendix	37

INTRODUCTION

I. *The Problem of Graduation:* In experimental work the results very frequently consist of a tabulated set of corresponding pairs of values of two variables arranged according to the ascending or descending sequence of one of the variables. For example,

$$\begin{array}{ccccccc} x_1 & x_2 & x_3 & x_4 & \cdots & x_n \\ y_1 & y_2 & y_3 & y_4 & \cdots & y_n \end{array}$$

where the x 's are arranged in a sequence. Due to observational error or other causes the values of $y_1, y_2, \cdots y_n$, in general, are not the ideal values of y corresponding respectively to $x_1, x_2, \cdots x_n$, but

* Assistant Professor of Drawing and Descriptive Geometry, University of Minnesota, Minneapolis, Minn.

differ from these ideal values by certain variable amounts. The data when plotted as points in the xy plane do not lie on a smooth curve. If a table of differences is formed from the given values, namely, $y_2 - y_1$, $y_3 - y_2$, $\dots$ $y_n - y_{n-1}$, it is usually found that these differences are irregular and do not adapt themselves to methods of interpolation and other uses to which a table of differences is applied.

The problem of smoothing, or graduation, is to determine a new set of y 's, namely, Y_1 , Y_2 , $\dots$ Y_n , which differ slightly from the given values and which when plotted with the x 's as points in the xy plane lie on a smooth curve.

2. *Other Methods of Smoothing:* The methods of smoothing data now in use may roughly be divided into three groups:

- (a) Graphical methods,
- (b) Fitting data to an assumed law,
- (c) Summation methods.

The first as its name implies is a graphical method, the other two are analytical although the second is sometimes thought of in a graphical sense.

(a) *Sprague's Method:* This is a method of smoothing which was presented to the Institute of Actuaries by its President, T. B. Sprague, in April, 1886 (*Journal of the Institute of Actuaries*, Oct., 1886). It is the method which naturally comes to mind first, namely, that of plotting the given values as points in the xy plane, drawing a smooth curve, and reading off the graduated values. As much information as possible is obtained from the data concerning the curve, such as concavity, maxima, minima, points of inflection, etc., before the curve is actually drawn. The graduated values are read as accurately as possible and the last figure is estimated. These last estimated figures are then still further adjusted so as to make the table of differences as regular as possible. The objection to the method is that it does not give the degree of refinement usually desired in most problems of graduation.

(b) *Fitting data to an assumed law:* Probably the best known method in this group is the so called method of Least Squares. It consists in assuming the analytical relation which exists between the two variables involved and then determining the constants in this relation by imposing the condition that the sum of the squares of the deviations between the given data and the data represented by the assumed analytical expression is to be a minimum.

If the law connecting the two variables is known the method of Least Squares is very satisfactory. The only objection to the method in that case is the length of time required to make a graduation. However, if the law joining the variables is not known the method of least squares presents difficulties which in some cases are insurmountable.

"The Application of the Pearsonian Frequency Curves to the Graduation of a Mortality Experience," as presented by W. P. Elderton before the Institute of Actuaries in 1903, is another example of the fitting of data to an assumed law. Still another example of this type of graduation is "The Application of the Logistic Function to Experimental Data" by Joseph Berkson and L. J. Reed, *Journal of Physical Chemistry*, May 1929.

If the law connecting the variables is not known, a rather serious objection to all the methods in this group is that the graduation may be too smooth; that is, the graduation may cover up certain information contained in the data by too great a smoothing.

(c) Summation Formulae: This type of graduation is nicely illustrated by the Woolhouse Formula of Graduation. It is briefly as follows: Suppose given fifteen points in the xy plane, uniformly spaced with respect to x . Since a parabola of 2nd degree is determined by three conditions, such a parabola may be passed through the first, sixth, and eleventh points. Similarly a second parabola may be passed through the second, seventh, and twelfth points; a third parabola through the third, eighth and thirteenth points; a fourth through the fourth, ninth, and fourteenth points; and a fifth parabola through the fifth, tenth, and fifteenth points. From the equation of each of these five parabolas may be determined the ordinate corresponding to the eighth point. The arithmetical mean of these five ordinates is taken as the graduated value of the ordinate for the eighth point. It is apparent that the graduated value of the ordinate for the eighth point is expressible by a formula involving the given quantities and hence the reason for the word "Formula" in the title. Another formula of this type is "Spencer's 21 Term Formula."

Such formulae require a large amount of data and they also require that one of the variables be in arithmetical progression. These requirements are not always possible in experimental work. The greatest disadvantage in their use is that the graduation does not cover the entire range of the data.

In addition to the methods of graduation mentioned above there are several others which do not fall naturally into any particular group. It is not intended, however, in this introduction to discuss all the methods but rather to indicate briefly what has been done on the problem of graduation by mentioning a few of the more prominent groupings.

3. *The Principle of Areas*: On June 22, 1929, Prof. T. R. Running of the University of Michigan presented a paper before a conference on chemical education in which he described a new method of graduating data.¹ This method is unique in that it is based on the principle of areas, a principle which had not been used in any of the other methods of graduation.

The principle of areas is simply that the area under the curve $y = \phi'(x)$ over the interval $x_i \leq x \leq x_j$ is equal numerically to the ordinate of the curve $y = \phi(x)$ at $x = x_j$ minus the ordinate of the same curve at $x = x_i$, or stated symbolically,

$$\int_{x_i}^{x_j} \phi'(x) dx = \phi(x_j) - \phi(x_i),$$

where

$$\phi'(x) = \frac{d\phi(x)}{dx}.$$

Professor Running's method of graduation is much simpler than any of the methods previously discussed and gives results which compare favorably with any of them.

4. *The Method of Graduation Presented in This Paper*: The method of graduation presented in this paper is also based on the principle of areas. It may be considered as an extension or refinement of Professor Running's method in which certain features which were thought to be objectionable in his method have been eliminated. Although the calculations are considerably longer than those in Professor Running's method, they are simple and the total amount of time necessary to make a graduation is much less than that required in the other methods which have been mentioned.

The advantages of the method of graduation presented here are:

The calculations are simple.

The graduation is over the entire range.

The graduation of each value depends on all the data and not simply on the data in the vicinity of that particular value.

¹ "Mathematics for Chemical Engineers," *Transactions of the American Institute of Chemical Engineers*, Vol. XXIII, 1929.

It is unnecessary that the values of either variable be in arithmetical progression.

No assumed relation or law is required.

The graduation is not too smooth; that is, it does not cover up information.

The method is particularly adaptable to interpolation.

DESCRIPTION OF METHOD

5. *Definitions:* Certain terms which will be used frequently in this paper are defined as follows:

Smooth: The word *smooth* as applied to a curve has already been used with a great deal of freedom and it is necessary to know what this word implies. By a smooth curve is meant a curve which is not only continuous but which has a continuous slope (1st derived curve).

Graduating or Integral Curve: The curve on which lie the graduated values, or the curve from which the given values deviate due to observational or other errors.

First Derived Curve: The curve whose ordinates represent the corresponding slopes on the integral curve. If the equation $y = \phi(x)$ represents the integral curve then $y = \phi'(x)$ is the analytical expression for the first derived curve, where

$$\phi'(x) = \frac{d\phi(x)}{dx}.$$

Second Derived Curve: The curve whose ordinates represent the corresponding slopes on the first derived curve. The second derived curve has the same relation to the first derived curve as the latter has to the integral curve. If $y = \phi(x)$ represents the integral curve, then $y = \phi''(x)$ represents the second derived curve, where

$$\phi''(x) = \frac{d^2\phi(x)}{dx^2}.$$

Deviation: By *deviation* is meant the difference between the given value and the corresponding graduated value. In order to specify the sense of the deviation, as far as this paper is concerned, deviation will mean the graduated value minus the given value. That is, if the deviation is positive it will mean that the graduated value is larger than the given value, and if negative, the graduate value is less than the given value.

6. *Explanation of Method:* Let it be supposed for the time being that it is possible to approximate the second derived curve by a broken line made up of segments of straight lines. Since this broken line is continuous the slope of the first derived curve will be continuous. If the condition is imposed on the first derived curve that its segments, which are parabolas of the second degree, meet at the same values of x as the corresponding segments of the second derived curve do, then the first derived curve is smooth.² Another condition imposed on the position of the first derived curve is that the total area under this curve over the whole range of data $x_1 \leq x \leq x_n$ is equal numerically to the quantity $y_n - y_1$.³ These conditions are sufficient to establish the analytical expressions for the segments of the first derived curve. If it be possible then to determine one ordinate on the integral or graduating curve all the other ordinates may be obtained by adding the proper increments of area to this particular ordinate. Inasmuch as the integral curve is built up by increments of area it will be continuous; and due to the smoothness of the first derived curve it will have a slope which is continuous. Hence the graduating curve will be a smooth curve composed of segments of parabolas of the third order which join not only with the same slope but with the same curvature.

The above paragraph expresses the essentials of the method in a few words. It remains to show how the graduation can be accomplished from the data. It is evident from what has been said, however, that the method is not purely graphical or purely analytical but is a combination of the two. Hence this method together with that of Professor Running does not fall into any of the groups mentioned in the introduction to this paper.

7. *Details of the Method:* It will first be shown how it is possible from the data to approximate the second derived curve as a broken line. Suppose there is given the data of columns 1 and 2 of chart I arranged according to the ascending sequence of the x 's.

² Here is the first major improvement of this method over that of Professor Running. While very little is assumed about the relationship between the two variables it is felt that the law which connects them when plotted as a curve does have at least a smooth first derived curve.

³ In Professor Running's method the total area under the first derived curve is $y_n - y_1$ but the area there is maintained piecemeal, that is, the area over the interval $x_1 \leq x \leq x_2$ is $y_2 - y_1$, the area over the interval $x_2 \leq x \leq x_3$ is $y_3 - y_2$, etc. A possible objection to this is that too much reliance is placed on individual readings.

1	2	3	4	5	6	7
x	y	Δx	Δy	$\frac{\Delta y}{\Delta x}$	Constructed Points	$\frac{\Delta^2 y}{\Delta x^2}$
x_1	y_1				L	
		$x_2 - x_1$	$y_2 - y_1$	$\frac{y_2 - y_1}{x_2 - x_1}$		$\frac{M - L}{x_2 - x_1}$
x_2	y_2				M	
		$x_3 - x_2$	$y_3 - y_2$	$\frac{y_3 - y_2}{x_3 - x_2}$		$\frac{N - M}{x_3 - x_2}$
x_3	y_3				N	
		$x_4 - x_3$	$y_4 - y_3$	$\frac{y_4 - y_3}{x_4 - x_3}$		$\frac{O - N}{x_4 - x_3}$
x_4	y_4				O	
		$x_5 - x_4$	$y_5 - y_4$	$\frac{y_5 - y_4}{x_5 - x_4}$		$\frac{P - O}{x_5 - x_4}$
x_5	y_5				P	

x_{n-1}	y_{n-1}				V	
		$x_n - x_{n-1}$	$y_n - y_{n-1}$	$\frac{y_n - y_{n-1}}{x_n - x_{n-1}}$		$\frac{W - V}{x_n - x_{n-1}}$
x_n	y_n				W	

CHART I

Column 3 is a table of the differences of the x 's, namely, $x_2 - x_1$, $x_3 - x_2$, $\dots$ $x_n - x_{n-1}$; similarly column 4 is a table of differences of the y 's and column 5 is a table of the differences of the y 's divided by the corresponding differences of the x 's, namely, $\frac{y_2 - y_1}{x_2 - x_1}$, $\frac{y_3 - y_2}{x_3 - x_2}$, $\dots$ $\frac{y_n - y_{n-1}}{x_n - x_{n-1}}$. The values of column 4, namely, $y_2 - y_1$, $y_3 - y_2$, $\dots$ $y_n - y_{n-1}$, are the areas under the ungraduated first derived curve in the respective intervals $x_1 \leq x \leq x_2$, $x_2 \leq x \leq x_3$, $\dots$ $x_{n-1} \leq x \leq x_n$. Hence the values of column 5, namely, $\frac{y_2 - y_1}{x_2 - x_1}$, $\frac{y_3 - y_2}{x_3 - x_2}$, $\dots$ $\frac{y_n - y_{n-1}}{x_n - x_{n-1}}$, are the heights of rectangles in these

same intervals whose areas are equal to the area under the first derived curve. These rectangles are indicated in figure 1.

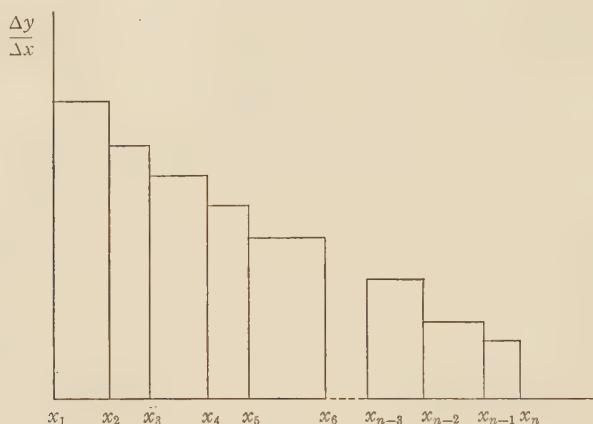


FIG. 1

Suppose in figure 2 that AB , CD , and EF are the ends of three adjacent rectangles shown in figure 1. Point G is the intersection of lines AD and BC . Point H is the intersection of lines CF and DE .

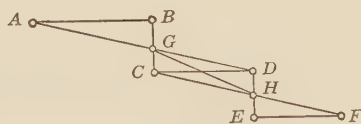


FIG. 2

If the intervals are equal the slope of line GH is the average slope of the first derived curve throughout the middle interval. If the intervals are not equal, that is, if the x 's are not in arithmetical progression, it is possible from the data to correct the slope of GH so as to obtain the average slope of the first derived curve throughout the interval.⁴ However, in most cases it will be found that this correction factor is negligible.

While the above construction is graphical it is possible from the data to determine analytically the slope of the line GH without drawing the figure. From the similar triangles ABG and DCG ,

$$\frac{BG}{BC} = \frac{AB}{AB + CD}.$$

⁴ See Article 9.

Knowing the ordinates of B and C , together with the size of the intervals AB and CD , the ordinate of point G can by the above expression be determined directly from the data. Similarly the ordinate of point H can be determined, and so on for the other corresponding points. Figure 3 is a reproduction of figure 1 in

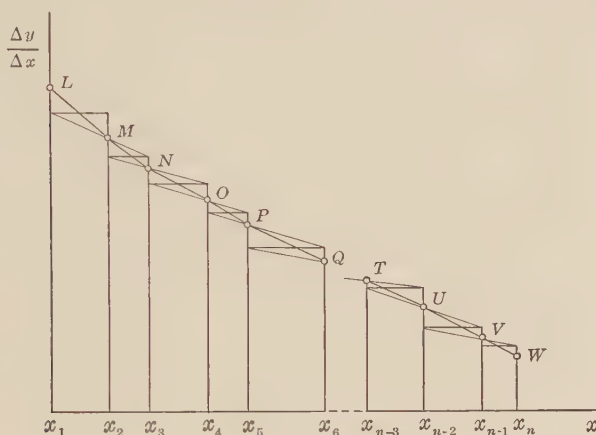


FIG. 3

which these so-called "constructed points" are shown. The slope of MN is the average slope of the first derived curve throughout the interval $x_2 \leq x \leq x_3$; the slope of NO is the average slope of the same curve over the interval $x_3 \leq x \leq x_4$; $\dots$; the slope of UV is the average slope of the same curve over the interval $x_{n-2} \leq x \leq x_{n-1}$. Hence it is possible to determine the average slope of the first derived curve by intervals from $x = x_2$ to $x = x_{n-1}$. In the end sections the lines LM and VW are drawn so as to bisect the ends of the rectangles. The slopes of LM and VW approximate then the average slope of the first derived curve in these end sections.⁵ The broken line shown in figure 3 is not the first derived curve but its segments have slopes which are the average slopes of the first derived curve throughout those intervals.

Column 6 of chart I is a table of the ordinates of the points $L, M, N, \dots, U, V, W$. In the table letter L represents the ordinate of point L in the figure, etc.

⁵ In certain problems from the nature of the quantities involved, readings beyond the range of the data can be estimated so that in such problems the construction described above will take care of the end sections as well as the interior sections.

The differences $M - L$, $N - M$, $O - N$, $P - O$, $Q - P$, $\dots$, $V - U$, $W - V$ are the areas under the second derived curve in the respective intervals $x_1 \leq x \leq x_2$, $x_2 \leq x \leq x_3$, $\dots$, $x_{n-1} \leq x \leq x_n$. Hence $\frac{M-L}{x_2-x_1}$, $\frac{N-M}{x_3-x_2}$, $\dots$, $\frac{W-V}{x_n-x_{n-1}}$ (column 7, chart 1) are the heights of rectangles in these same intervals whose areas are equal to the areas under the second derived curve. These rectangles are indicated in figure 4.

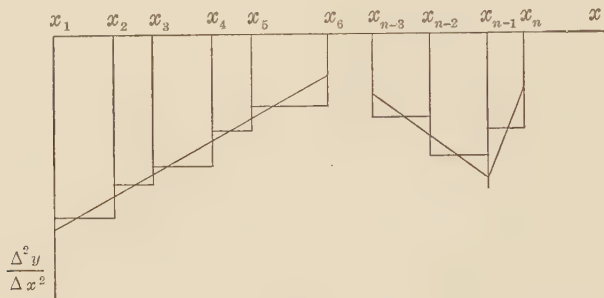


FIG. 4

The second derived curve is then approximated by drawing a broken line which follows the trend indicated by the rectangles and in such a way that the area under the broken line approximates the sum of the areas of the rectangles. If the area under the broken line differs appreciably from the sum of the areas of the rectangles the broken line is shifted in the y direction an amount equal to

$$\frac{\text{Sum of the Areas of Rectangles} - \text{Area Under Broken Line}}{x_n - x_1}$$

8. *Calculations:* If the second derived curve is a broken line composed of h straight line segments the equations of these segments may be written:

$$\begin{array}{llll} y = 2m_ax + q_a & \text{over the interval} & x_1 \leq x \leq x_a \\ y = 2m_bx + q_b & \text{“ “ “} & x_a \leq x \leq x_b \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ y = 2m_hx + q_h & \text{“ “ “} & x_{h-1} \leq x \leq x_n \end{array}$$

The equations of the corresponding segments of the first derived curve are:

$$y = m_a x^2 + q_a x + c_a$$

$$y = m_b x^2 + q_b x + c_b$$

$$\cdot$$

$$\cdot$$

$$y = m_h x^2 + q_h x + c_h$$

where $c_a, c_b, c_c, \dots, c_h$ are constants of integration whose values are to be determined.

If the restriction is placed on the segments of the first derived curve that the first two meet at $x = x_a$, that the second and third meet at $x = x_b$; etc., there results the following relations between the constants of integration:

$$c_b - c_a = x_a^2 (m_a - m_b) + x_a (q_a - q_b) \quad (1)$$

$$c_c - c_b = x_b^2 (m_b - m_c) + x_b (q_b - q_c) \quad (2)$$

$$\cdot$$

$$\cdot$$

$$c_h - c_{h-1} = x_{h-1}^2 (m_{h-1} - m_h) + x_{h-1} (q_{h-1} - q_h) \quad (h - 1)$$

Hence there are $h - 1$ simultaneous equations in h unknowns.

A more convenient set of equations (1)', (2)', $\dots$ (h - 1)' is obtained from this set by repeating the first equation for the first equation of the new set, by adding the first and second equations of this set for the second equation of the new set, by adding the first three equations of this set for the third equation of the new set, etc.

The new set of equations may then be written

$$c_b = c_a + K_a \quad (1)'$$

$$c_c = c_a + K_b \quad (2)'$$

$$\cdot$$

$$\cdot$$

$$\cdot$$

$$c_h = c_a + K_{h-1} \quad (h - 1)'$$

where K_a represents the right hand member of (1); K_b represents the sum of the right hand members of (1) and (2); K_c represents the sum of the right hand members of (1), (2) and (3); etc. An additional condition is imposed on the constants $c_a, c_b, \dots, c_h$ by requiring that the area under the first derived curve over the entire range

$x_1 \leq x \leq x_n$ is equal numerically to the value $y_n - y_1$. This condition is expressed by the equation:

$$y_n - y_1 = \int_{x_1}^{x_a} (m_a x^2 + q_a x + c_a) dx + \int_{x_a}^{x_b} (m_b x^2 + q_b x + c_b) dx \cdots \\ + \int_{x_{h-1}}^{x_n} (m_h x^2 + q_h x + c_h) dx$$

which reduces to the form

$$(x_a - x_1)c_a + (x_b - x_a)c_b \cdots + (x_n - x_{h-1})c_h = W \quad (h)$$

If the values of $c_b, c_c, \cdots c_h$ as given by (1)', (2)', $\cdots (h-1)'$ respectively are inserted in (h) the value of c_a is determined. With this value of c_a the values of $c_b, c_c, \cdots c_h$ are then given by (1)', (2)', $\cdots (h-1)'$, respectively.

If $Y_1, Y_2, Y_3, \cdots Y_n$ are the graduated values of y (not yet determined) corresponding respectively to the given values of $y_1, y_2, \cdots y_n$, then the area under the first derived curve over the in-

1	2	3	4	5
x	y	Areas	Sums	Y
x_1	y_1			Y_1
		$Y_2 - Y_1$	$Y_2 - Y_1$	
x_2	y_2			Y_2
		$Y_3 - Y_2$	$Y_3 - Y_1$	
x_3	y_3			Y_3
		$Y_4 - Y_3$	$Y_4 - Y_1$	
x_4	y_4			Y_4
		$Y_5 - Y_4$	$Y_5 - Y_1$	
x_5	y_5			Y_5
		$Y_6 - Y_5$	$Y_6 - Y_1$	
x_{n-1}	y_{n-1}			Y_{n-1}
		$Y_n - Y_{n-1}$	$Y_n - Y_1$	
x_n	y_n			Y_n

CHART 2

terval $x_1 \leq x \leq x_2$ is equal numerically to $Y_2 - Y_1$; the area under the same curve over the interval $x_2 \leq x \leq x_3$, is equal numerically to $Y_3 - Y_2$; etc. In chart 2 as in chart 1 columns 1 and 2 are the given data. The areas under the first derived curve over any interval are determined from the analytical expressions for the segments of that curve. They are represented in chart 2, column 3, by their equivalents $Y_2 - Y_1$, $Y_3 - Y_2$, etc. Column 4 gives the sums of the areas of column 3; the first entry is $Y_2 - Y_1$; the second entry is $(Y_3 - Y_2) + (Y_2 - Y_1)$ or $Y_3 - Y_1$; etc. The sum of column 2 is

$$y_1 + y_2 + y_3 + \cdots + y_{n-1} + y_n.$$

The sum of column 4 is

$$Y_2 + Y_3 + Y_4 + \cdots + Y_{n-1} + Y_n - (n - 1)Y_1.$$

If the sum of column 4 is subtracted from the sum of column 2 and this difference represented by G , then

$$G = y_1 + y_2 + y_3 + \cdots + y_{n-1} + y_n - Y_2 - Y_3 - Y_4 - \cdots - Y_{n-1} - Y_n + (n - 1)Y_1.$$

Now if

$$y_1 + y_2 + y_3 + y_4 + \cdots + y_n = Y_1 + Y_2 + Y_3 + Y_4 + \cdots + Y_n,$$

then

$$G = nY_1.$$

Hence the initial graduated value Y_1 is

$$Y_1 = \frac{G}{n}$$

By adding to Y_1 the first entry of column 3 the value of Y_2 is obtained. Adding to Y_2 the second entry of column 3 the value of Y_3 is obtained, and in a similar way all the other graduated values of y .

9. *Correction Factor:* The construction described in article 7 and illustrated by figure 2 was used to give the average slope of the first derived curve throughout an interval. In general the error in this assumption is small and is neglected. However, an expression for the necessary correction factor will be developed here in case it seems desirable in some particular problem to use it.

In figure 5 as in figure 2 are shown the extremities of three adjacent rectangles in the figure for the first derived curve. Here, for convenience, points will be designated by numbers.

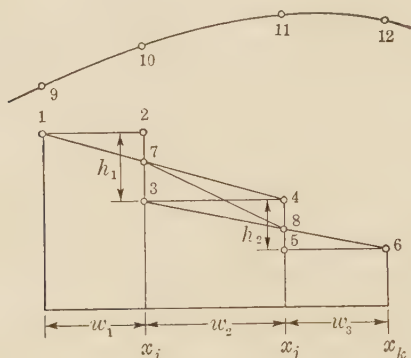


FIG. 5

The assumption is that the slope of line 7-8 is the average slope of the first derived curve throughout the interval $x_i \leq x \leq x_j$.

Let it be assumed that the three intervals considered are in an interval covered by one segment of the integral curve which is represented in the figure by the curve 9-10-11-12. There is no loss in generality as far as this discussion is concerned in assuming point 9 as the origin of the system of coordinates.

If the equation of curve 9-10-11-12 is

$$y = Ax + Bx^2 + Cx^3$$

the coordinates of the four points 9, 10, 11, 12 are:

$$\begin{aligned} &9(0, 0), \\ &10(x_i, Ax_i + Bx_i^2 + Cx_i^3), \\ &11(x_j, Ax_j + Bx_j^2 + Cx_j^3), \\ &12(x_k, Ax_k + Bx_k^2 + Cx_k^3). \end{aligned}$$

From the tabulated results shown on page 15 the coordinates of points 1, 2, 3, 4, 5, and 6 are:

$$\begin{aligned} &1(0, A + Bx_i + Cx_i^2), \\ &2(x_i, A + Bx_i + Cx_i^2), \\ &3(x_i, A + B(x_j + x_i) + C(x_j^2 + x_jx_i + x_i^2)), \\ &4(x_j, A + B(x_j + x_i) + C(x_j^2 + x_jx_i + x_i^2)), \\ &5(x_j, A + B(x_k + x_j) + C(x_k^2 + x_kx_j + x_j^2)), \\ &6(x_k, A + B(x_k + x_j) + C(x_k^2 + x_kx_j + x_j^2)). \end{aligned}$$

Point	x	y	Δx	Δy	$\frac{\Delta y}{\Delta x}$
9	0	0			
			x_i	$Ax_i + Bx_i^2 + Cx_i^3$	$A + Bx_i + Cx_i^2$
10	x_i	$Ax_i + Bx_i^2 + Cx_i^3$			
			$x_j - x_i$	$A(x_j - x_i) + B(x_j^2 - x_i^2) + C(x_j^3 - x_i^3)$	$A + B(x_i + x_j) + C(x_i^2 + x_j^2 + x_i x_j)$
11	x_j	$Ax_j + Bx_j^2 + Cx_j^3$			
			$x_k - x_j$	$A(x_k - x_j) + B(x_k^2 - x_j^2) + C(x_k^3 - x_j^3)$	$A + B(x_k + x_j) + C(x_k^2 + x_j^2 + x_k x_j)$
12	x_k	$Ax_k + Bx_k^2 + Cx_k^3$			

The equation of line 1-4 is

$$y = x[B + C(x_j + x_i)] + A + Bx_i + Cx_i^2.$$

The equation of line 3-6 is

$$y = x[B + C(x_i + x_j + x_k)] + A + Bx_j + C(x_j^2 - x_ix_k).$$

Point 7 is the intersection of line 1-4 and line 2-3. Point 8 is the intersection of line 3-6 and line 4-5. The coordinates of these two points are:

$$7[x_i, A + 2Bx_i + C(2x_i^2 + x_ix_j)],$$

$$8[x_j, A + 2Bx_j + C(2x_j^2 + x_jx_k + x_ix_j - x_ix_k)].$$

The slope of line 7-8 is

$$2B + C(2x_j + 2x_i + x_k). \quad (1)$$

Suppose points 13 and 14 to be two points on the first derived curve having the respective x coordinates x_i and x_j .

Since $y = Ax + Bx^2 + Cx^3$ is the equation of the integral curve, the equation of the first derived curve is

$$y = A + 2Bx + 3Cx^2.$$

The coordinates of points 13 and 14 are then:

$$13(x_i, A + 2Bx_i + 3Cx_i^2),$$

$$14(x_j, A + 2Bx_j + 3Cx_j^2).$$

The average slope of the first derived curve over the interval $x_i \leq x \leq x_j$ is the slope of chord 13-14 or

$$2B + 3C(x_j + x_i).^6 \quad (2)$$

⁶ The average slope of a parabola of second degree between two of its points is the slope of the chord joining the two points.

Proof: Let the equation of the parabola be



$$y = ax^2 + bx + c.$$

Let the coordinates of the two points be $P(e, f)$, $Q(g, h)$. The slope of the parabola

The assumption made at the beginning of this article is then that expression (1) is equal to expression (2).

If E represents the percentage of error in the assumption, then

$$E = \frac{[2B + 3C(x_j + x_i) - 2B - C(2x_j + 2x_i + x_k)] \cdot 100}{2B + 3C(x_j + x_i)} \\ = \frac{C(x_j + x_i - x_k) \cdot 100}{2B + 3C(x_j + x_i)}. \quad (3)$$

Let the widths of the three intervals considered be represented by w_1 , w_2 , and w_3 respectively; that is,

$$w_1 = x_i, \\ w_2 = x_j - x_i, \\ w_3 = x_k - x_j, \quad \text{where} \quad x_i < x_j < x_k.$$

The expression for E may then be written:

$$E = \frac{C(w_1 - w_3)}{2B + 3C(2w_1 + w_2)} \cdot 100. \quad (4)$$

Equation 4 shows that the assumption made is correct if the x 's are in arithmetical progression, since in that case $w_1 = w_3$ and $E = 0$.

Equation 4 is still in a form which cannot conveniently be used as a correction factor.

If h_1 represents the ordinate of point 3 (figure 5) minus the ordinate of point 2 and similarly if h_2 represents the ordinate of point 5 minus the ordinate of point 4, then

$$h_1 = Bx_j + C(x_j^2 + x_jx_i) \\ = x_j[B + C(x_j + x_i)] \\ = (w_1 + w_2)[B + C(2w_1 + w_2)] \quad (5)$$

and

$$h_2 = B(x_k - x_i) + C(x_k^2 + x_kx_j - x_jx_i - x_i^2) \\ = (x_k - x_i)[B + C(x_i + x_j + x_k)] \\ = (w_2 + w_3)[B + C(3w_1 + 2w_2 + w_3)]. \quad (6)$$

at any point is $2ax + b$. The average slope of the parabola from P to Q is

$$\frac{\int_e^g (2ax + b) dx}{\int_e^g dx} = \frac{[ax^2 + bx]_e^g}{g - e} = a(g + e) + b.$$

The slope of chord PQ is

$$\frac{h - f}{g - e} = \frac{ag^2 + bg + c - (ae^2 + be + c)}{g - e} = a(g + e) + b. \quad \text{Q.E.D.}$$

Equation (5) may be written

$$B = \frac{h_1}{w_1 + w_2} - C(2w_1 + w_2). \quad (5)'$$

Substituting this value of B in (6) there results

$$C = \frac{\frac{h_2}{w_2 + w_3} - \frac{h_1}{w_1 + w_2}}{w_1 + w_2 + w_3}$$

or

$$C = \frac{h_2(w_1 + w_2) - h_1(w_2 + w_3)}{(w_1 + w_2)(w_1 + w_2 + w_3)(w_2 + w_3)}. \quad (7)$$

Solving (5) and (6) in a similar way for B there results

$$B = \frac{h_1(w_3 + 2w_2 + 3w_1)(w_2 + w_3) - h_2(w_1 + w_2)(2w_1 + w_2)}{(w_1 + w_2)(w_1 + w_2 + w_3)(w_2 + w_3)}. \quad (8)$$

If now the values of C and B as given in (7) and (8) are substituted in (4) there results

$$E = \frac{(w_1 - w_3)[h_2(w_1 + w_2) - h_1(w_2 + w_3)] \cdot 100}{h_1[(w_2 + w_3)(2w_3 + w_2)] + h_2[(w_1 + w_2)(2w_1 + w_2)]}. \quad (9)$$

As an illustration of the application of this formula suppose $h_1 = h_2$ and $w_1 = w_2 = 2w_3$; then

$$E = \frac{w_3 h_1 (4w_3 - 3w_3) 100}{h_1 [(3w_3)(4w_3) + (4w_3)(6w_3)]} = \frac{h_1 w_3^2 100}{h_1 w_3^2 36} = 2.8.$$

The assumed value is 97.2 per cent of the correct value.

Another illustration of the application of this formula is the following:

Suppose $h_1 = 2h_2$, $w_1 = 2w_3$, $w_2 = 4w_3$; then

$$E = \frac{100 w_3 h_2 (6w_3 - 10w_3)}{h_2 [2(5w_3)(6w_3) + (6w_3)(8w_3)]} = \frac{100 h_2 w_3^2 (-4)}{h_2 w_3^2 (108)} = -3.7.$$

The assumed value is 103.7 per cent of the correct value.

ILLUSTRATIVE EXAMPLES

For the sake of brevity, only two illustrative examples will be given. The first example nicely illustrates the method of solution.

The second problem is one which has caused considerable difficulty in solving by other methods of smoothing. In addition to illustrating the method, this example demonstrates how certain difficulties are taken care of which arise in particular problems.

10. *A Cooling Vessel of Water:* This example is illustrative of a large number of problems which are encountered in laboratory work. It offers no particular difficulties and for that reason is convenient to use as a first illustration of the method of graduation.

The observed temperatures (y) in degrees centigrade of a vessel of cooling water at time (x) in minutes from the beginning of observation are given in the following table.⁷

x	0	1	2	3	5	7	10	15	20
y	92.0	85.3	79.5	74.5	67.0	60.5	53.5	45.0	39.5

(a) *Solution*

The given data are tabulated in columns 1 and 2 of chart 3. Columns 3 and 4 are the differences of x and y respectively. Column 5 gives the divided differences $\Delta y/\Delta x$, or the heights of the rectangles whose areas equal the areas under the first derived curve. Column 6 gives the ordinates of the so-called "constructed points" which are used to determine the average slope of the first derived curve throughout the various intervals. Column 7 gives these average slopes of the first derived curve (designated $\Delta^2 y/\Delta x^2$), or the heights of the rectangles whose areas equal the areas under the second derived curve. Rectangles with heights equal to the quantities in column 7 are shown in figure 6. The broken line $PQRST$ is drawn to approximate the second derived curve. Column 8 of chart 3 gives the ordinates of these five points:

$P(0, .92)$

$Q(3, .79)$

$R(5, .36)$

$S(10, .17)$

$T(20, .09)$

The area under the broken line $PQRST$ is 6.34. The area of the combined rectangles is the difference of the extreme values of column

⁷ Lipka: "Graphical and Mechanical Computation," page 259.

CHART 3

1	2	3	4	5	6	7	8	9	10	11
x	y	Δx	Δy	$\frac{\Delta y}{\Delta x}$	Constr. Points	$\frac{\Delta^2 y}{\Delta x^2}$	Pts. on 2nd Curve	Areas	Sums	Y
0	92.0				-7.15		.92			92.074
		1	-6.7	-6.7		.90		-6.683	-6.683	
1	85.3				-6.25					85.391
		1	-5.8	-5.8		.85		-5.807	-12.490	
2	79.5				-5.40					79.584
		1	-5.0	-5.0		.82		-4.973	-17.463	
3	74.5				-4.58		.79			74.611
		2	-7.5	-3.75		.54		-7.848	-25.311	
5	67.0				-3.50		.36			66.763
		2	-6.5	-3.25		.31		-6.172	-31.483	
7	60.5				-2.88					60.591
		3	-7.0	-2.33		.26		-7.224	-38.707	
10	53.5				-2.09		.17			53.367
		5	-8.5	-1.70		.14		-8.522	-47.229	
15	45.0				-1.40					44.845
		5	-5.5	-1.10		.12		-5.272	-52.501	
20	39.5				-.80		.09			39.573
	596.8								-231.867	596.799

6, chart 3, namely $-.80 + 7.15$, or 6.35 . The correction of the position of the line $PQRST$ is

$$\frac{6.35 - 6.34}{20} = \frac{.01}{20} = .0005,$$

which is negligible.⁸

In the interval $0 \leq x \leq 3$:

The equation of the second derived curve (line PQ) is $y = -.0433x + .92$.

⁸ If this correction were used the coordinates of the five points would be $P(0, .9205)$, $Q(3, .7905)$, $R(5, .3605)$, $S(10, .1705)$, and $T(20, .0905)$.

The equation of the first derived curve is

$$y = -.02167x^2 + .92x + C_1. \quad (1)$$

In the interval $3 \leq x \leq 5$:

The equation of the second derived curve (line QR) is
 $y = -.215x + 1.435$.

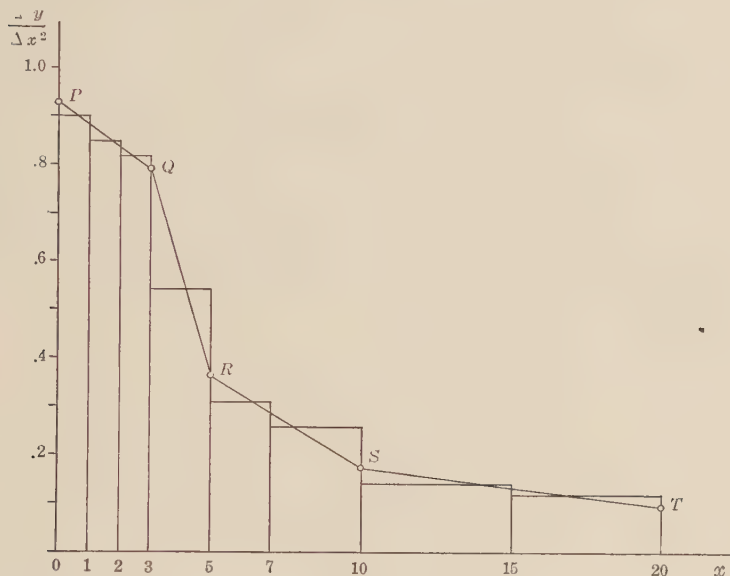


FIG. 6

The equation of the first derived curve is

$$y = -.1075x^2 + 1.435x + C_2. \quad (2)$$

In the interval $5 \leq x \leq 10$:

The equation of the second derived curve (line RS) is
 $y = -.038x + .55$.

The equation of the first derived curve is

$$y = -.019x^2 + .55x + C_3. \quad (3)$$

In the interval $10 \leq x \leq 20$:

The equation of the second derived curve (line ST) is
 $y = -.008x + .25$.

The equation of the first derived curve is

$$y = -.004x^2 + .25x + C_4. \quad (4)$$

The conditions that curves (1) and (2) meet at $x = 3$, curves (2) and (3) meet at $x = 5$, and curves (3) and (4) meet at $x = 10$ are respectively:⁹

$$C_2 - C_1 = \frac{1}{2} \times 3(.92 - 1.435) = -.7725, \quad (5)$$

$$C_3 - C_2 = \frac{1}{2} \times 5(1.435 - .55) = 2.2125, \quad (6)$$

$$C_4 - C_3 = \frac{1}{2} \times 10(.55 - .25) = 1.5000. \quad (7)$$

Equation (5) may be written

$$C_2 = C_1 - .7725. \quad (5)'$$

$$\text{Adding (5) and (6),} \quad C_3 = C_1 + 1.4400. \quad (6)'$$

$$\text{Adding (5), (6), and (7),} \quad C_4 = C_1 + 2.9400. \quad (7)'$$

The area under the first derived curve from $x = 0$ to $x = 20$ is the difference of the extreme values of y , namely, $39.5 - 92.0 = -52.5$.

Therefore

$$\begin{aligned} -52.5 &= \int_0^3 (-.02167x^2 + .92x + C_1)dx \\ &\quad + \int_3^5 (-.1075x^2 + 1.435x + C_2)dx \\ &\quad + \int_5^{10} (-.019x^2 + .55x + C_3)dx \\ &\quad + \int_{10}^{20} (-.004x^2 + .25x + C_4)dx \end{aligned}$$

⁹ If $y = Ax^2 + Bx + C_1$ and $y = Dx^2 + Ex + C_2$ are the equations of two segments of the first derived curve the condition that they meet at $x = m$ is $C_2 - C_1 = \frac{1}{2} \cdot m(B - E)$.

Proof: The corresponding segments of the second derived curve are $y = 2Ax + B$ and $y = 2Dx + E$. Under the condition of the method these meet at $x = m$; therefore

$$2Am + B = 2Dm + E$$

or

$$m(A - D) = -\frac{1}{2}(B - E). \quad (1)$$

The condition that the two segments of the first derived curve meet at $x = m$ is

$$Am^2 + Bm + C_1 = Dm^2 + Em + C_2$$

or

$$C_2 - C_1 = (A - D)m^2 + (B - E)m. \quad (2)$$

But from (1)

$$(A - D)m^2 = -\frac{1}{2}m(B - E). \quad (3)$$

Substituting (3) in (2) there results

$$C_2 - C_1 = \frac{1}{2}m(B - E). \quad \text{Q.E.D.}$$

or

$$3C_1 + 2C_2 + 5C_3 + 10C_4 = -107.6633. \quad (8)$$

Inserting the values of C_2 , C_3 , and C_4 as given by (5)', (6)', and (7)', respectively, in (8), there results

$$20C_1 = -142.7183$$

or

$$C_1 = -7.136. \quad (9)$$

Therefore

$$C_2 = -7.908, \quad (10)$$

$$C_3 = -5.696, \quad (11)$$

$$C_4 = -4.196. \quad (12)$$

As a check on the calculations:

The area under the first derived curve

from	$x = 0$	to	$x = 3$	is	-17.463
"	$x = 3$	to	$x = 5$	is	-7.848
"	$x = 5$	to	$x = 10$	is	-13.397
"	$x = 10$	to	$x = 20$	is	-13.793

or the total area is

$$\underline{-52.501}$$

The equations of the segments of the first derived curve, then, are:

Interval	$0 \leq x \leq 3$	$y = -.02167x^2 + .92x - 7.136$
"	$3 \leq x \leq 5$	$y = -.1075x^2 + 1.435x - 7.908$
"	$5 \leq x \leq 10$	$y = -.019x^2 + .55x - 5.696$
"	$10 \leq x \leq 20$	$y = -.004x^2 + .25x - 4.196$

Column 9 of chart 3 gives the areas under the first derived curve throughout the respective intervals. Column 10, chart 3, gives the progressive sums of these areas.

The sum of column 2 is 596.8. The sum of column 10 is -231.867 . The initial graduated value of y is

$$Y_1 = \frac{596.8 + 231.867}{9} = 92.074.$$

The second graduated value is $92.074 - 6.683$ or 85.391 ; the third graduated value is $85.391 - 5.807$ or 79.584 ; etc.

The graduated values may also be obtained by adding each value of column 10 to the initial graduated value 92.074. For example, the value of Y_3 is $92.074 - 12.490 = 79.584$.

The graduated data is shown in column 11, chart 3.

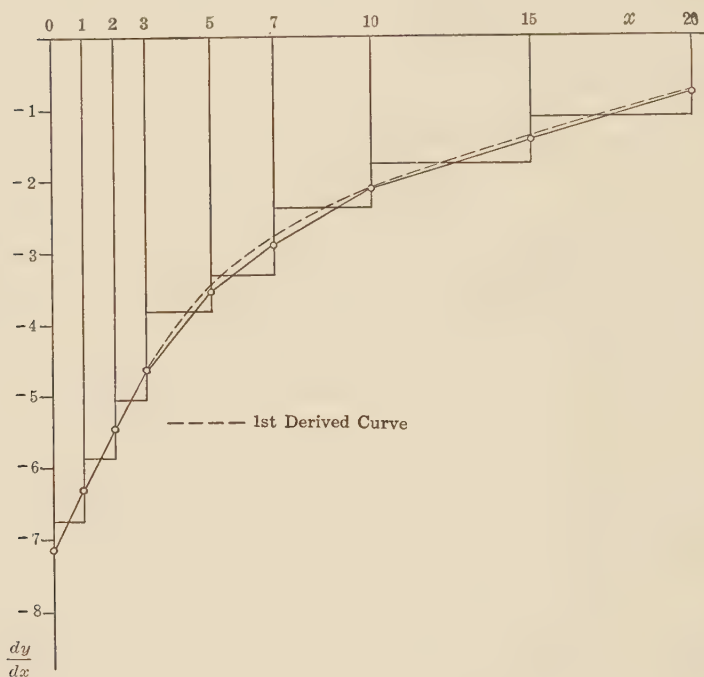


FIG. 7

In figure 7 are shown the rectangles (height $\Delta y/\Delta x$), the constructed points, the average slopes of the first derived curve and the first derived curve itself.

(b) Comparison with the Method of Least Squares

The results of this graduation are compared with those obtained by the method of least squares in chart 4. Columns 3 and 6 give the graduations by this method and by that of Least Squares, respectively.¹⁰ Columns 4 and 7 give the respective deviations. Columns 5 and 8 give the squares of these deviations respectively.

¹⁰ The graduation by the method of least squares was made on the assumption that the data obeyed the law $y = Ae^{kx} + B$.

CHART 4

1	2	3	4	5	6	7	8
Data		Method of this Paper			Method of Least Squares		
x	y	Y	Dev.	Dev. ²	Y	Dev.	Dev. ²
0	92.0	92.074	+ .074	.005476	91.367	-.633	.400689
1	85.3	85.391	+ .091	.008281	85.375	+ .075	.005625
2	79.5	79.584	+ .084	.007056	80.005	+ .505	.255025
3	74.5	74.611	+ .111	.012321	75.195	+ .695	.483025
5	67.0	66.763	-.237	.056169	67.022	+ .022	.000484
7	60.5	60.591	+ .091	.008281	60.462	-.038	.001444
10	53.5	53.367	-.133	.017689	52.967	-.533	.284089
15	45.0	44.845	-.155	.024025	44.852	-.148	.021904
20	39.5	39.573	+ .073	.005329	40.166	+ .666	.443556
	596.8	596.799		.144627	597.411		1.695841

Criteria of a good graduation are that the sum of the graduated values should equal the sum of given values, that the sum of the squares of the deviations should be small, and that the number of reversals in the sign of the deviations should be large. It will be noted that the graduation given here excels that given by the method of least squares in all three respects.

(c) *Interpolation*

In the introduction mention was made of the fact that the method of graduation described here adapts itself readily to interpolation. In the above illustrated example the areas under the first derived curve over any interval whatever can be easily obtained after the analytical expressions for the segments have once been established and the initial graduated value determined. Chart 5 shows the given data in columns 1 and 2. Column 3 gives the calculated areas under the first derived curve for the intervals indicated, and column 4 gives the graduated values of y for all integer values of x from $x = 0$ to $x = 20$.

CHART 5

1	2	3	4
x	y	Areas	Y
0	92.0		92.074
		-6.683	
1	85.3		85.391
		-5.807	
2	79.5		79.584
		-4.973	
3	74.5		74.611
		-4.211	
4			70.400
		-3.637	
5	67.0		66.763
		-3.247	
6			63.516
		-2.925	
7	60.5		60.591
		-2.642	
8			57.949
		-2.395	
9			55.554
		-2.187	
10	53.5		53.367

CHART 5—Continued

1	2	3	4
x	y	Areas	Y
		-2.013	
11			51.354
		-1.850	
12			49.504
		-1.696	
13			47.808
		-1.551	
14			46.257
		-1.412	
15	45.0		44.845
		-1.282	
16			43.563
		-1.161	
17			42.402
		-1.046	
18			41.356
		-.940	
19			40.416
		-.843	
20	39.5		39.573

II. *Reaction of Diphenyl-chlor-methan with Alcohol*

Diphenyl-chlor-methan reacts in ethyl alcohol, yielding hydrochloric acid and an ester. In an experiment the following data were obtained where "y" represents the mols HCl (also mols ester) at time " θ ." ¹¹

θ	y
1	.001520
3	.001979
5	.002455
8	.003098
10	.00367
16	.00484
21	.00586
26	.00693
31	.00795
41	.00994
51	.01190
61	.01387
63	.01430
103	.02135
125	.02492
151	.02855
175	.03246
1037	.07179

The chlor compound at the start was 0.08808 normal. The experiment was allowed to proceed until y remained constant at 0.07405 for two days in succession.

(a) *Pseudo Readings*

A difficulty which presents itself at once in the solution of this graduation problem is that ninety-four per cent of the data covers only seventeen per cent of the range.

¹¹ Hitchcock and Robinson: "Differential Equations in Applied Chemistry."

If the data is plotted on uniform coordinate paper the distance in the θ direction between the last two readings is 862 units. However, if it be plotted on a logarithmic scale the distance between the last two readings in the θ direction is .77 of a unit (logarithmic deck). Applying Sprague's method ¹² then to the data but using a logarithmic scale in the θ direction pseudo readings may be obtained over the range $175 \leq \theta \leq 1037$.

If the data is plotted on semi-log paper (uniform scale for y , logarithmic scale for θ) the last four readings lie approximately on a straight line. On log-log paper all the values seem to lie on a curve.

In chart 6, column 1 gives the value of " θ "; column 2 gives the value of y read from the log-log curve; column 3 gives the values of y read from the semi-log curve and column 4 gives the assumed values or the so-called Pseudo-Readings.¹³

CHART 6

1	2	3	4
θ	y log-log	y semi-log	y assumed
300	.0458	.0438	.045
400	.0535	.0503	.052
500	.0593	.0553	.0575
600	.0635	.0592	.0615
700	.0665	.0625	.065
800	.0685	.0655	.0675
900	.0700	.0683	.070

In the graduation these pseudo readings will be used only to indicate the behavior of the function from $\theta = 175$ to $\theta = 1037$, and they will not be used in calculating the initial graduated value of Y .

(b) Smoothing out of rough data: Columns 1 and 2 of chart 7 tabulate the given data. Column 3 of the same chart lists the values

¹² See article 2(a) of the Introduction.

¹³ Not the mean values of columns 2 and 3, but rough values in the vicinity of these two.

of $\Delta y/\Delta\theta$. An inspection of this column indicates a more or less erratic fluctuation in values. From the nature of the experiment it is known that the first derived curve is concave upward throughout the entire range. The fluctuations in the values of column 3 are not due then to any inherent property of the function y but rather due to error caused by the difficulty in taking readings. Individual fluctuations in this graduation will therefore be smoothed out. It will be noted that the values of $\Delta y/\Delta\theta$ from $\theta = 1$ to $\theta = 63$ fluctuate around .000206, the average value of $\Delta y/\Delta\theta$ over that interval.

Instead of considering individual values of $\Delta y/\Delta\theta$ the average value .000206 will be taken over interval $1 \leq \theta \leq 63$ and the average value .000154 over the interval $103 \leq \theta \leq 175$.

(c) *Solution*

The ordinates of the "constructed points" are listed in column 4 of chart 7.

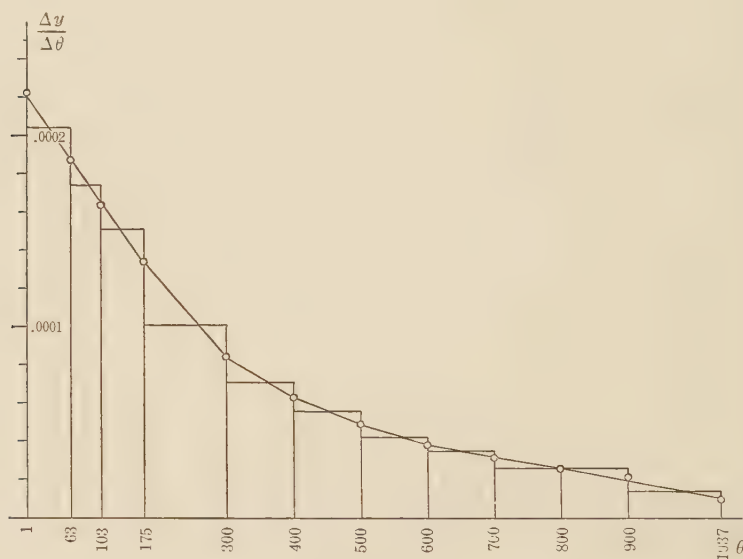


FIG. 8

The rectangles and the average slopes of the first derived curve are represented graphically in figure 8. The average slopes are so nearly alined from $\theta = 1$ to $\theta = 175$ and from $\theta = 600$ to $\theta = 1037$ that they are taken as such over those intervals.

CHART 7

1	2	3	4	5	6	7	8	9
θ	y	$\frac{\Delta y}{\Delta \theta}$	Constructed Points	$\frac{\Delta^2 y}{\Delta \theta^2}$	Points on 2nd Curve	Area	Sums	Y
1	.001520		.000224		-5.17K			.001485
		.000230				.000444	.000444	
3	.001979							.001929
		.000238				.000441	.000885	
5	.002455							.002370
		.000214				.000658	.001543	
8	.003098							.003028
		.000286				.000436	.001979	
10	.00367							.003464
		.000195				.001296	.003276	
16	.00484							.004760
		.000204				.001066	.004342	
21	.00586			-5.17K				.005826
		.000214				.001053	.005395	
26	.00693							.006879
		.000204				.001040	.006435	
31	.00795							.007919

CHART 7—Continued

1	2	3	4	5	6	7	8	9
θ	y	$\frac{\Delta y}{\Delta \theta}$	Constructed Points	$\frac{\Delta^2 y}{\Delta \theta^2}$	Points on 2nd Curve	Area	Sums	Y
41	.00994	.000199				.002042	.008477	.009961
51	.01190	.000196				.001990	.010467	.011951
61	.01387	.000197				.001938	.012405	.013889
63	.01430	.000215				.000382	.012787	
103	.02135	.000176	.000188	-5.17K		.007195	.019981	.014271
125	.02492	.000162	.000168		-5.17K	.003606	.023587	.021466
151	.02855	.000140		-5.17K		.003946	.027533	.029018
175	.03246	.000163	.000134			.003349	.030882	.032367
					-4.77K			

CHART 7—Continued

1	2	3	4	5	6	7	8	9
θ	y	$\frac{\Delta y}{\Delta \theta}$	Constructed Points	$\frac{\Delta^2 y}{\Delta \theta^2}$	Points on 2nd Curve	Area	Sums	Y
		.000100		-4.08K		.013473		
300	.045		.000083		-2.92K			.045840
		.000070		-2.00K		.007303		
400	.052		.000063		-1.75K			.053143
		.000055		-1.50K		.005434		
500	.0575		.000048		-1.25K			.058577
		.000040		-1.00K		.004137		
600	.0615		.000038		-.843K			.062714
		.000035		-.686K		.003245		
700	.065		.000030		-.686K			.065959
		.000025		-.686K		.002533		
800	.0675		.000025					.068492
		.000025		-.686K		.001847		
900	.070		.000018					.070339
		.000013		-.686K		.001416	.070270	
1037	.07179		.000008		-.686K			.071755

 $K = 10^{-7}$

Column 5, chart 7, lists the values of the average slopes of the first derived curve or the heights of rectangles in the figure for the second derived curve. The letter " K " in this column represents 10^{-7} . To avoid confusion the letter " K " will be used throughout the rest of the problem to represent this quantity.

Rectangles with heights equal to the values listed in column 5 are represented graphically in figure 9.

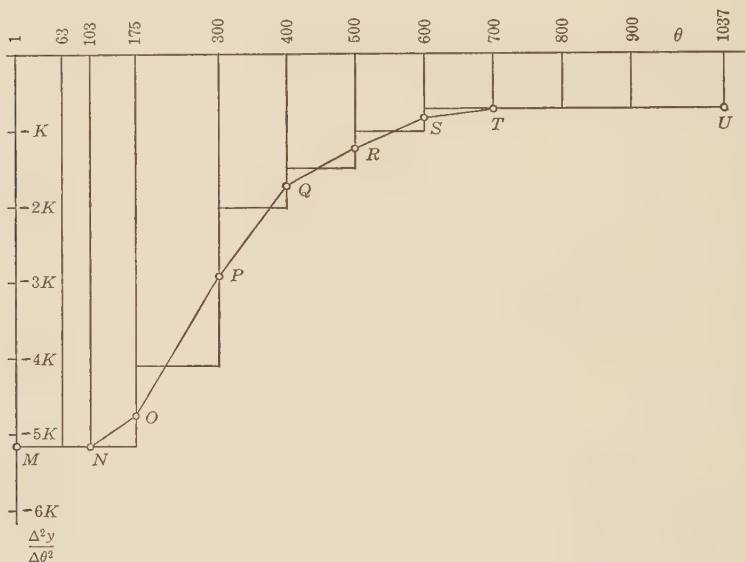


FIG. 9

The second derived curve is the broken line $MNOPQRSTU$ of Fig. 9.¹⁴ The ordinates of these 9 points are listed in column 6 of chart 7.

The combined area of the rectangles, Fig. 9, is $-2160.0K$.

The area under the broken line is $-2166.24K$.

The correction in the position of the broken line is

$$\left[\frac{-2160.00 + 2166.24}{1036} \right] K \quad \text{or} \quad +.006K$$

which is negligible.

¹⁴ The points $M, N, O, P, Q, R, S, T, U$, were obtained by the same construction as that used to obtain the "constructed points" of Fig. 8. In problems where the second derived curve is fairly regular this establishes an exact method of drawing the second derived curve.

The equations of the segments of the second and first derived curves respectively over the various intervals are:

Interval $1 \leq \theta \leq 103$

Second Derived Curve $y = -5.17K$
(line *MN*)

First Derived Curve $y = (-5.17\theta + C_1)K$

Interval $103 \leq \theta \leq 175$

Second Derived Curve $y = (.005555\theta - 5.7422)K$
(line *NO*)

First Derived Curve $y = (.002778\theta^2 - 5.7422\theta + C_2)K$

Interval $175 \leq \theta \leq 300$

Second Derived Curve $y = (.0148\theta - 7.36)K$
(line *OP*)

First Derived Curve $y = (.0074\theta^2 - 7.36\theta + C_3)K$

Interval $300 \leq \theta \leq 400$

Second Derived Curve $y = (.0117\theta - 6.43)K$
(line *PQ*)

First Derived Curve $y = (.00585\theta^2 - 6.43\theta + C_4)K$

Interval $400 \leq \theta \leq 600$

Second Derived Curve $y = (.004535\theta - 3.564)K$
(line *QS*)

First Derived Curve $y = (.002268\theta^2 - 3.564\theta + C_5)K$

Interval $600 \leq \theta \leq 700$

Second Derived Curve $y = (.00157\theta - 1.785)K$
(line *ST*)

First Derived Curve $y = (.000785\theta^2 - 1.785\theta + C_6)K$

Interval $700 \leq \theta \leq 1037$

Second Derived Curve $y = -.686K$
(line *TU*)

First Derived Curve $y = (-.686\theta + C_7)K$

The conditions that the first and second segments of the first derived curve meet at $\theta = 103$, that the second and third segments meet at $\theta = 175$, etc., are:

$$C_2 - C_1 = \frac{1}{2}(-5.17 + 5.7422)103 = 29.47$$

$$C_3 - C_2 = \frac{1}{2}(-5.7422 + 7.36)175 = 141.56$$

$$C_4 - C_3 = \frac{1}{2}(-7.36 + 6.43)300 = -139.50$$

$$C_5 - C_4 = \frac{1}{2}(-6.43 + 3.564)400 = -573.20$$

$$C_6 - C_5 = \frac{1}{2}(-3.564 + 1.785)600 = -533.70$$

$$C_7 - C_6 = \frac{1}{2}(-1.785 + .686)700 = -384.65$$

The progressive sums of the above equations are:

$$C_2 = C_1 + 29.47 \quad (1)$$

$$C_3 = C_1 + 171.03 \quad (2)$$

$$C_4 = C_1 + 31.53 \quad (3)$$

$$C_5 = C_1 - 541.67 \quad (4)$$

$$C_6 = C_1 - 1075.37 \quad (5)$$

$$C_7 = C_1 - 1460.02 \quad (6)$$

The area under the first derived curve from $\theta = 1$ to $\theta = 1037$ is .07179 - .00152 or .07027; therefore

$$\begin{aligned} \frac{.07027}{K} = 702,700 &= \int_1^{103} (-5.17\theta + C_1)d\theta \\ &+ \int_{103}^{175} (.002778\theta^2 - 5.7422\theta + C_2)d\theta \\ &+ \int_{175}^{300} (.0074\theta^2 - 7.36\theta + C_3)d\theta \\ &+ \int_{300}^{400} (.00585\theta^2 - 6.43\theta + C_4)d\theta \\ &+ \int_{400}^{600} (.002268\theta^2 - 3.564\theta + C_5)d\theta \\ &+ \int_{600}^{700} (.000785\theta^2 - 1.785\theta + C_6)d\theta \\ &+ \int_{700}^{1037} (-.686\theta + C_7)d\theta \end{aligned}$$

or

$$\begin{aligned} 102C_1 + 72C_2 + 125C_3 + 100C_4 + 200C_5 + 100C_6 \\ + 337C_7 = 1,626,747 \quad (7) \end{aligned}$$

Solving equations (1), (2), (3), (4), (5), (6), and (7) simultaneously there results

$$C_1 = 2227.79 \quad (8)$$

$$C_2 = 2257.26 \quad (9)$$

$$C_3 = 2398.82 \quad (10)$$

$$C_4 = 2259.32 \quad (11)$$

$$C_5 = 1686.12 \quad (12)$$

$$C_6 = 1152.42 \quad (13)$$

$$C_7 = 767.77 \quad (14)$$

With these constants established, the area under the first derived curve throughout any interval can be calculated. These areas are listed in column 7 of chart 7 and their progressive sums are listed in column 8 of the same chart.¹⁵

The sum of the given values of y is .267412. The sum of column 8 is .240688. The initial graduated value of y is

$$Y_1 = \frac{.267412 - .240688}{18} = .001485.$$

The second graduated value is obtained by adding the first entry of column 7 to Y_1 ; the third graduated value is obtained by adding the second entry of column 7 to Y_2 ; etc.

The graduated values of y are listed in column 9 of chart 7.

APPENDIX

(A) *Use of Graduation to Determine Constants in Differential Equation*

If the reaction of diphenyl-chlor-methan with ethyl alcohol (article 11) is assumed to be of the form



then the following differential equation

$$\frac{dy}{d\theta} = K_1(.08808 - y) - K_2y^2 \quad (1)$$

expresses the relation between y and θ , where K_1 and K_2 are constants.

¹⁵ The pseudo values are not used in calculating the initial graduated value.

The results of the graduation of article 11 may be used to determine the constants K_1 and K_2 of equation (1) as well as to determine whether this equation does express the conditions of the reaction or not.

The quantity $dy/d\theta$ is the slope of the graduating curve or the ordinate of the first derived curve. If a graduated value of y and the corresponding ordinate of the first derived curve are inserted in equation (1) for y and $dy/d\theta$, respectively there will result an equation involving K_1 and K_2 for each given value of θ . If these eighteen equations are divided into two groups and if then the equations in each group are combined by addition into a single equation there results two equations which can be solved for K_1 and K_2 .

The constants obtained in this manner are:

$$K_1 = .002574,$$

$$K_2 = .006526.^{16}$$

If the above values of K_1 and K_2 are inserted in equation (1) this equation gives a value of $dy/d\theta$ for each graduated value of y . If the values of $dy/d\theta$ obtained in this manner agree reasonably well with the corresponding ones obtained from the analytical expressions for the first derived curve (see article 11), then it is probable that the given differential equation does express the reaction and that the values of K_1 and K_2 which have been used are approximately correct.

In chart 8, column 1 gives the values of θ ; column 2 gives the values of $dy/d\theta$ obtained from the analytical expressions for the first derived curve (article 11); and column 3 gives the values of $dy/d\theta$ obtained from the differential equation using the graduated values of y and the above values of K_1 and K_2 . The figures in chart 8, while not agreeing exactly, are close enough to indicate that the assumed differential equation does express the reaction.

(B) To Verify the Fact That the Second Derived Curve of Article 11 Does Have One Point of Inflection

It was found in article 11 that the second derived curve had one point of inflection (see figure 9). By making use of the differential equation (1) of appendix A as well as the values of K_1 and K_2

¹⁶ These values were obtained by grouping the first nine equations together and the second nine together. If alternate equations are grouped together slightly different values are obtained, namely, $K_1 = .002575$ and $K_2 = .006619$.

CHART 8

1	2	3
θ	$\frac{dy}{d\theta}$	$\frac{dy}{d\theta}$
1	.0002223	.0002229
3	.0002212	.0002217
5	.0002202	.0002206
8	.0002186	.0002189
10	.0002176	.0002177
16	.0002145	.0002143
21	.0002119	.0002115
26	.0002093	.0002087
31	.0002068	.0002059
41	.0002016	.0002004
51	.0001964	.0001950
61	.0001912	.0001897
63	.0001902	.0001887
103	.0001695	.0001685
125	.0001583	.0001581
151	.0001454	.0001465
175	.0001337	.0001366
1037	.0000057	.0000084
∞	0 ¹⁷	.0000003

determined there it is possible not only to confirm the fact that the second derived curve has one point of inflection but also to determine its position.

The equation of the reaction is assumed to be

$$y' = .08808K_1 - K_1y - K_2y^2 \quad \text{where} \quad y' = \frac{dy}{d\theta}. \quad (1)$$

¹⁷ The value $\frac{dy}{d\theta} = 0$ for $\theta = \infty$ is assumed. In determining the corresponding quantity of column 3 the given value $y = .07405$ was used.

Differentiating with respect to θ

$$\begin{aligned} y'' &= -K_1 y' - 2K_2 y y' \\ &= -y'(K_1 + 2K_2 y). \end{aligned} \quad (2)$$

Differentiating a second time

$$\begin{aligned} y''' &= -y'(2K_2 y') + (K_1 + 2K_2 y)(-y'') \\ &= -y''(K_1 + 2K_2 y) - 2K_2 (y')^2. \end{aligned} \quad (3)$$

Differentiating again

$$\begin{aligned} y'''' &= -y''(2K_2 y') + (K_1 + 2K_2 y)(-y''') - 4K_2 y' y'' \\ &= -y'''(K_1 + 2K_2 y) - 6K_2 y' y''. \end{aligned} \quad (4)$$

If the second derived curve has a point of inflection the value of y'''' at that point is zero. Setting y'''' equal to zero in (4), there results:

$$-y'''(K_1 + 2K_2 y) - 6K_2 y' y'' = 0. \quad (5)$$

If the value of y''' as given by (3) is substituted in (5) the latter equation becomes:

$$y''(K_1 + 2K_2 y)^2 + 2K_2 (y')^2 (K_1 + 2K_2 y) - 6K_2 y' y'' = 0. \quad (6)$$

Substituting in the last term of (6) the value of y'' as given by (2), equation (6) becomes:

$$(K_1 + 2K_2 y)[y''(K_1 + 2K_2 y) + 8K_2 (y')^2] = 0. \quad (7)$$

The factor $K_1 + 2K_2 y$ cannot be zero since K_1 and K_2 are positive constants and y is positive over the entire range under consideration. This factor may accordingly be factored out of (7) leaving:

$$y''(K_1 + 2K_2 y) + 8K_2 (y')^2 = 0. \quad (8)$$

Inserting in (8) the value of y'' as given by (2), there results

$$-y'(K_1 + 2K_2 y)^2 + 8K_2 (y')^2 = 0. \quad (9)$$

The quantity y' is not equal to zero within the range considered and it may be factored out of (9), leaving:

$$-(K_1 + 2K_2 y)^2 + 8K_2 y' = 0. \quad (10)$$

If the value for y' as given by the equation (1) be inserted in (10) there results

$$-(K_1 + 2K_2 y)^2 + 8K_2 (.08808K_1 - K_1 y - K_2 y^2) = 0$$

or better

$$y^2 + \frac{K_1}{K_2}y + \frac{K_1^2 - 8(.08808)K_1K_2}{12K_2^2} = 0. \quad (11)$$

Solving (11) for y there is obtained:

$$y = \frac{-K_1 \pm \sqrt{\frac{2}{3}K_1^2 + .23488K_1K_2}}{2K_2}. \quad (12)$$

Since positive values of y are the only ones of interest it is apparent from (12) that the second derived curve has one point of inflection, provided

$$.23488K_2 > 1/3K_1.$$

The above inequality is satisfied by the values of K_1 and K_2 found in appendix A, and hence there is one point of inflection on the second derived curve.

The position of this point is determined by inserting in (12) the values of K_1 and K_2 . The value of y thus found is .02435.

$\therefore$ The second derived curve has one point of inflection situated at $y = .02435$ or better (by interpolation) at $\theta = 121$.

If the right hand member of (3) be set equal to zero it can be shown in a similar way that the first derived curve has no point of inflection.

